ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2005
Seria: TRANSPORT z. 58 Nr kol. 1688

Grzegorz KAROŃ[1], Renata ŻOCHOWSKA[2]

ZASTOSOWANIE STRUMIENIOWYCH BAZ DANYCH W SYSTEMACH TRANSPORTOWYCH

Streszczenie. Tematem artykułu jest propozycja wykorzystania systemów strumieniowych baz danych DSMS w analizie potoków ruchu. Przedstawiono w nim koncepcję rejestrowania strumienia danych opisujących potoki ruchu w sieci transportowej oraz zaproponowano schemat strumienia danych na potrzeby analizy ruchu w sieci ulic. Przetwarzanie danych z wykorzystaniem komend języka CQL zrealizowano w systemie DSMS: STREAM opracowanym na Uniwersytecie Stanford [15].

APPLICATION OF DATA STREAM SYSTEMS IN TRANSPORTATION SYSTEMS

Summary. The proposition of data stream systems DSMS application in traffic flows analysis is a subject of the article. The record of data stream that describes traffic flows in transportation network is a concept presented in the paper. The authors have also proposed the scheme of data stream for traffic analysis in roadnets. Data processing with application of CQL has been realised in DSMS system: STREAM developed in Stanford University.

1. WPROWADZENIE

Sieć dróg (ulic) jest częścią systemu transportowego umożliwiającą jego funkcjonowanie, to znaczy przemieszczanie osób i towarów pomiędzy punktami (węzłami) nadania a punktami (węzłami) odbioru. Rozkład potoków ruchu w sieci zależy między innymi od wielkości tych potoków, macierzy podróży, wielkości i gęstości sieci drogowej, elementów sterowania ruchem (znaki drogowe, sygnalizacja, strefy ograniczonego ruchu, strefy ograniczonego parkowania itp.). Sprawna organizacja ruchu w sieci, której celem jest zapewnienie odpowiedniej płynności ruchu oraz właściwego poziomu swobody ruchu, to jedno z głównych zadań inżynierii ruchu. Niezbędnym zagadnieniem jest zbieranie i analiza danych opisujących potoki pojazdów.

Współczesny rozwój metod detekcji oraz rejestracji, w tym wideodetekcja, umożliwiają w miarę łatwe i szybkie pozyskiwanie informacji dotyczących miedzy innymi prędkości i rodzaju pojazdów. Dodatkowo nastąpił szybki rozwój systemów nawigacji satelitarnej opartych na systemie określania pozycji obiektów GPS (ang. *Global Positioning System)* oraz

[1] Wydział Transportu, Politechnika Śląska, ul. Krasińskiego 8, 40-019 Katowice, tel/fax (+48 32) 2554159, grzegorz.karon@polsl.pl
[2] Wydział Transportu, Politechnika Śląska, ul. Krasińskiego 8, 40-019 Katowice, tel/fax (+48 32) 2554121, renzoc@polsl.katowice.pl

systemie przesyłania informacji GSM (ang. *Global System for Mobile Communications*). Dzięki tym systemom informacje o aktualnym rozkładzie potoków w sieci przekazywane są do urządzeń sterowania ruchem oraz do kierowców, co powoduje zmiany rozkładu potoków ruchu uwzględniające aktualną sytuację ruchową w sieci (przeciążenia odcinków ulic, występowanie korków, ograniczenia w ruchu wywołane awariami infrastruktury lub kolizjami pojazdów).

Wzrost natężenia ruchu powoduje wzrost ilości informacji rejestrowanych w czasie przez systemy. Problemem staje się gromadzenie tych informacji w bazach danych oraz szybkość i efektywność przetwarzania na dane użyteczne dla systemów sterowania ruchem, zwłaszcza przetwarzanie danych w czasie rzeczywistym (on-line), umożliwiające szybką reakcję na zakłócenia w ruchu [12].

W następnych rozdziałach opracowania przedstawiono koncepcję przetwarzania danych napływających w postaci strumieni do strumieniowej bazy danych DSMS. Wykorzystano prototypowy system STREAM dostępny na serwerach Uniwersytetu Stanford [15].

2. REJESTRACJA STRUMIENIA POJAZDÓW

Na rysunku 1 przedstawiono fragment sieci ulic centrum Katowic. Do analizy strumienia danych wybrano ulicę Francuską, na której znajduje się 7 ponumerowanych kolejno skrzyżowań. Odcinki międzywęzłowe oznaczono liczbami utworzonymi z numerów kolejnych węzłów (skrzyżowań). Założono, że na odcinkach znajdują się urządzenia rejestrujące przejazd samochodów (na przykład rejestratory wykorzystujące rozpoznawanie obrazu) oraz następujące parametry: numer rejestracyjny (część cyfrową), rodzaj pojazdu (SOD, SC, AP, itd.), prędkość pojazdu (wyliczaną na podstawie odległości pomiędzy rejestratorami oraz czasu, jaki dzieli chwile rejestracji pojazdu w dwóch kolejnych rejestratorach). Dane z poszczególnych rejestratorów odcinkowych tworzą strumień danych o schemacie: moment zgłoszenia, numer identyfikacyjny, rodzaj pojazdu, prędkość pojazdu, numer odcinka, nazwa odcinka, numer węzła początkowego odcinka, numer węzła końcowego odcinka.

Rys. 1. Fragment sieci transportowej Katowic z zaznaczonymi węzłami ulicy Francuskiej o numerach 1, 2, 3, 4, 5, 6, 7 oraz odcinkami międzywęzłowymi o numerach: 12, 23, 34 i 67

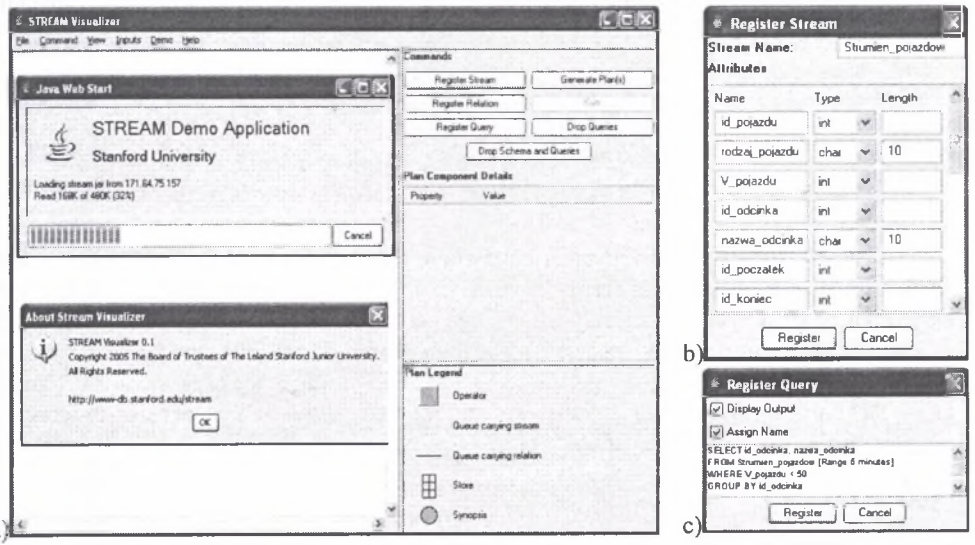
Fig. 1. Road network of Katowice City with location of intersections: 1, 2, 3, 4, 5, 6, 7 and links (roads): 12, 23, 34, 67

3. SYSTEM STRUMIENIOWEJ BAZY DANYCH STREAM

System zarządzania strumieniową bazą danych STREAM [6,15](ang. *STanford stReam datA Manager*) wykorzystujący język CQL [1,2,5](ang. *Continous Query Language*) opracowany został w Stanach Zjednoczonych na Uniwersytecie Stanford. System posiada architekturę klient-serwer przystosowaną do pracy w środowisku WWW (ang. *Word Wide Web*). Elementem systemu komunikującym się z użytkownikiem jest klient bazy danych STREAM Visualiser pokazany na rys.2a. Jest to aplikacja napisana w języku Java, instalowana przez Internet na komputerze użytkownika, której zadaniem jest prezentacja danych oraz przekazywanie danych do serwera bazy danych STREAM, pracującego na komputerach uniwersyteckich (Host IP: 171.64.75.157 Port 12005 oraz IP: 171.64.75.156 Port 12001).

Serwer STREAM pracuje w dwóch fazach. W fazie pierwszej rejestruje przesyłane przez klienta: strumienie danych, relacje oraz zapytania w języku CQL. Rejestracja polega na sprawdzeniu poprawności danych ze schematami opisującymi strumienie, relacje i zapytania oraz na ich optymalizacji podnoszącej sprawność przetwarzania bazy danych przez system DSMS (ang. *Data Stream Management System*). W drugiej fazie uruchamiany jest strumień danych, na którym wykonywane są zarejestrowane relacje i zapytania [15].

Rejestrowanie strumienia (ang. *register stream*) przypomina tworzenie tabeli w relacyjnej bazie danych i polega na nadaniu mu nazwy (nazwy tabeli) oraz zdefiniowaniu jego atrybutów (pól tabeli). Na rysunku 2b pokazano rejestrację strumienia `Strumien_pojazdów`, którego opis zawiera tabl. 1.



Rys. 2. Aplikacja STREAM – System zarządzania strumieniowymi bazami danych DSMS: a) główne okno aplikacji, b) rejestrowanie w systemie strumienia, c) rejestrowanie w systemie zapytania w języku CQL

Fig. 2. DSMS STREAM – STanford stReam datA Manager: a) main window of application, b) data stream registration, c) CQL query registration

Tablica 1

Opis atrybutów strumienia wejściowego: `Strumien_pojazdow`

Lp.	Nazwa pola	Znaczenie	Typ
1	`timestamp`	czas pojawienia się danych w strumieniu - to pole występuje w strumieniu wejściowym, jednak nie jest definiowane przez użytkownika	całkowity
2	`id_pojazdu`	identyfikator pojazdu	całkowity
3	`rodzaj_pojazdu`	osobowy, ciężarowy, autobus, itd.	tekstowy
4	`V_pojazdu`	prędkość chwilowa pojazdu	całkowity
5	`id_odcinka`	identyfikator odcinka międzywęzłowego (ulicy)	całkowity
6	`nazwa_odcinka`	nazwa odcinka międzywęzłowego (ulicy)	tekstowy
7	`id_poczatek`	identyfikator węzła początkowego odcinka	całkowity
8	`id_koniec`	identyfikator węzła końcowego odcinka	całkowity

Źródło: opracowanie własne

Rejestrowanie zapytania (ang. *register query*) polega na zapisaniu zapytania w języku CQL i przesłaniu go do serwera, który sprawdzi składnię oraz wygeneruje plan wykonania zapytania.

Na rysunku 2c pokazano sposób rejestrowania zapytania, które tworzy relację zawierającą zestawienie odcinków międzywęzłowych, na których w ciągu ostatnich 5 minut pojazdy poruszały się z prędkością mniejszą niż 50 km/h – są to potencjalne odcinki, na których poziom swobody ruchu jest ograniczony. Poniżej przedstawiono treść zapytania CQL:

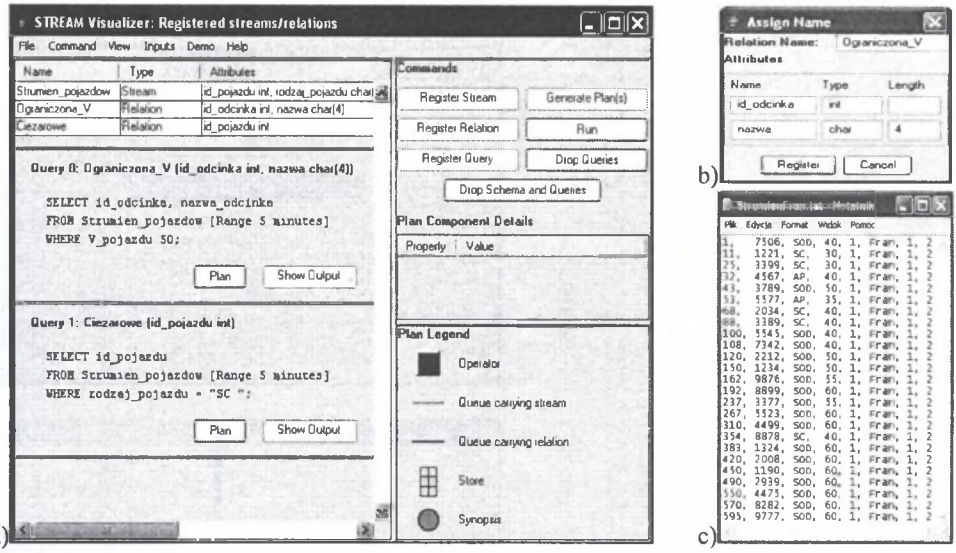
```
Query0:

SELECT id_odcinka, nazwa_odcinka FROM Strumien_pojazdow [Range 5
minutes]
WHERE V_pojazdu < 50
```

Wyniki powyższego zapytania zestawione zostaną w tablicy (relacji), którą należy zdefiniować przez nadanie jej nazwy `Ograniczona_V` oraz zdefiniowaniu pól – w tym przypadku będą to pola: `id_odcinka` i `nazwa` (rys.3b). Podobnie odbywa się rejestracja zapytania CQL, które tworzy relację zawierającą numery identyfikacyjne (`id_pojazdu`) pojazdów ciężarowych (`rodzaj_pojazdu = "SC "` – uwaga: w cudzysłowiu są 3 znaki: litery SC oraz jedna spacja, ponieważ atrybut jest typu znakowego o długości 4 pomniejszonej o 1 – specyfika języka Java i C), które przejechały w ciągu ostatnich 5 minut:

```
Query1:

SELECT id_pojazdu FROM Strumien_pojazdow [Range 5 minutes]
WHERE rodzaj_pojazdu = "SC "
```



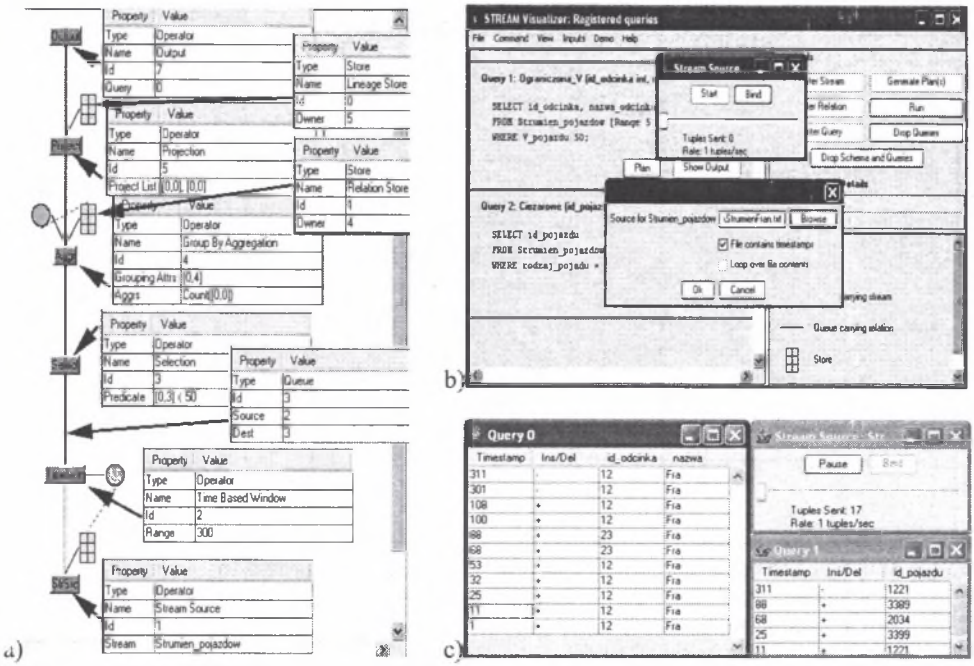
Rys. 3. Aplikacja STREAM – System zarządzania strumieniowymi bazami danych DSMS: a) strumień, relacje i zapytania zarejestrowane w systemie, b) rejestrowanie relacji, c) plik zawierający dane strumieniowe

Fig. 3. DSMS STREAM – STanford stReam datA Manager: a) stream, relations and queries registered in system, b) relation registration, c) input data file with data stream

Plan DSMS jest etapem przygotowania systemu STREAM do pracy, w którym powstają plany przedstawiające kolejność wykonywania elementów zapytań. Na rysunku 4a przedstawiono plan zapytania `Query0`. Plan przedstawiony jest w postaci schematu blokowego, na który składają się następujące bloki:

- StrSrc – operator przedstawiający źródło strumienia danych,
- TimeWin – okno czasowe zawierające zakres analizowanych danych strumienia,
- Select – operator wybierania elementów strumienia według zdefiniowanych w zapytaniu warunków,
- Aggr – operator grupowania elementów według określonego pola,
- Project – operator tworzenia widoku – tabeli z wynikami zapytania,
- Output – operator przedstawiający wyjście strumienia danych,
- Store – operator przechowywania danych w pamięci.

Każdy blok opisany jest poprzez właściwości (ang. *property*) oraz ich wartości (ang. *value*). Wymienione bloki połączone są liniami symbolizującymi kolejki (ang. *queue*) danych powstające podczas realizacji poszczególnych operacji elementarnych - operatorów.



Rys. 4. Aplikacja STREAM – System zarządzania strumieniowymi bazami danych DSMS: a) plan zapytania CQL, b) rejestrowanie pliku źródłowego, c) relacje wynikowe przetwarzania zarejestrowanych zapytań

Fig. 4. DSMS STREAM – STanford stReam datA Manager: a) CQL query plan, b) input data file registration, c) output relations

Źródło danych jest ostatnim elementem przygotowania systemu do przetwarzania danych. System STREAM pobiera dane z pliku tekstowego, w którym w kolejnych wierszach powinny być zapisane poszczególne krotki strumienia z atrybutami oddzielonymi przecinkami. Pierwszym atrybutem musi być moment pojawienia się krotki w strumieniu (ang. *timestamp*), a kolejne atrybuty muszą odpowiadać zdefiniowanemu schematowi strumienia.

Dla omawianego przykładu schemat strumienia (kolejne atrybuty) zestawiono w tabl. 1, zawartość pliku tekstowego z przykładowymi danymi pokazano na rys. 3c, natomiast rejestrację pliku źródłowego przedstawiono na rys. 4b.

Uruchomienie systemu (ang. *run*) powoduje rozpoczęcie wysyłania krotek z pliku źródłowego w chwilach czasowych zgodnych z pierwszym atrybutem (ang. *timestamp*) oraz wykonywanie zdefiniowanych zapytań i tworzenie widoków (tabel) z wynikowymi relacjami. Rysunek 4c przedstawia relację zapytań `Query0`, `Query1` oraz informację o 17 wysłanych do systemu krotkach. Pole Ins/Del informuje, czy krotka została dodana (+) czy usunięta (-) z relacji w miarę przesuwania się strumienia przez zdefiniowane w zapytaniach okno czasowe o szerokości 5 minut (`[Range 5 minutes]`).

Więcej przykładowych zapytań użytecznych dla systemu przetwarzania danych opisujących potoki ruchu pojazdów zawiera opracowanie [12].

4. PODSUMOWANIE

W artykule przedstawiono propozycję systemu rejestracji i przetwarzania strumienia danych opisujących ruch pojazdów w sieci transportowej. Wykorzystano nowy model danych: dane strumieniowe oraz nowy system bazy danych: system zarządzania strumieniową bazą danych DSMS. Systemy DSMS są obecnie rozwijane i stają się coraz popularniejsze, ponieważ mają istotną przewagę nad systemami relacyjnych baz danych RDBMS - działają na danych dostarczanych on-line w postaci strumieni krotek rejestrowanych w chwilach czasowych. Dzięki opracowanej na Uniwersytecie Stanford aplikacji STREAM możliwe było przetestowanie projektu systemu rejestracji potoków ruchu pojazdów.

Przewiduje się dalsze prace rozwojowe zmierzające do opracowania systemu wspomagającego organizację ruchu w sieci transportowej podczas dużego natężenia ruchu, występowania zakłóceń związanych z kolizjami lub nieplanowanymi wyłączeniami pasów ruchu.

Literatura

1. Arasu A., Babu S., Widom J.: The CQL Continous Query Language: Semantic foundations and query execution. Technical report, Stanford University Database Group, October 2003.
2. Arasu A., Widom J.: A Denotational Semantics for Continuous Queries over Streams and Relations. Technical Report, Mar. 2004.
3. Babcock B., Babu S., Datar M., Motwani R., Widom J.: Models and Issues in Data Stream Systems. Invited paper in Proc. of the 2002 ACM Symp. on Principles of Database Systems (PODS 2002), June 2002.
4. Babcock B., Datar M., Motwani R.: Load Shedding for Aggregation Queries over Data Streams. In Proc. of Intl. Conference on Data Engineering (ICDE 2004), March 2004.
5. Babu S., Widom J.: Continuous Queries over Data Streams. In SIGMOD Record, September 2001.
6. Babu S., Subramanian L., Widom J.: A Data Stream Management System for Network Traffic Management.
7. Beynon-Davies P.: Systemy baz danych. WNT, Warszawa 2000.
8. Chomicz P.: Hurtownie danych - współczesność, X Konferencja PLOUG Kościelisko, październik 2004.
9. Date C.: Wprowadzenie do baz danych. WNT, Warszawa 1981.
10. Date C. J., Darwen H.: SQL. Omówienie standardu języka. WNT, Warszawa 2000.
11. Garcia-Molina H. Ullman J. D., Widom J.: Implementacja systemów baz danych. WNT, Warszawa 2003.
12. Karoń G., Żochowska R.: Przetwarzanie informacji w systemach transportowych – strumieniowe bazy danych. III Konferencja naukowo-techniczna nt.: Systemy Transportowe – Teoria i Praktyka, Katowice 7 września 2005.
13. Kline K., Kline D.: SQL. Almanach. Opis poleceń języka. „Helion", Gliwice 2002.
14. Lausen G., Vossen G.: Obiektowe bazy danych. Modele danych i języki. WNT, Warszawa 2000.
15. Motwani R., Widom J. i inni: Query Processing, Resource Management, and Approximation in a Data Stream Management System. In Proc. of the 2003 Conf. on Innovative Data Systems Research (CIDR), January 2003.
16. Petersen J.: Wprowadzenie do baz danych. „Helion", Gliwice 2003.

17. Srivastava U., Munagala K., Widom J.: Operator Placement for In-Network Query Processing In Proc. of the 2004 ACM Symp. on Principles of Database Systems December 2004.
18. Srivastava U., Widom J.: Flexible time management in data stream systems. In Proc. of the 23rd ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems, pages 263-274, June 2004.
19. Thomas D., Motwani R.: Caching Queues in Memory Buffers In In Proc. of the Annual ACM-SIAM Symp. on Discrete Algorithms (SODA 2004), January 2004.
20. Todman Ch.: Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM). WNT, Warszawa 2003.
21. Tucker P.A., Maier D., Sheard T., Fegaras L.: Exploiting punctuation semantics in continous data stream. IEEE Trans. on Knowledge and Data Engg. 2003.
22. Ullman J. D., Widom J.: Podstawowy wykład z systemów baz danych. WNT, Warszawa 2001.